RATIONAL INATTENTION: A RESEARCH AGENDA

CHRISTOPHER A. SIMS

ABSTRACT. The literature applying information-theoretic ideas to economics has so far considered only Gaussian uncertainty. Ex post Gaussian uncertainty can be justified as optimal when the associated optimization problem is linear-quadratic, but the literature has often assumed Gaussian uncertainty even where it cannot be justified as optimal. This paper considers a simple two-period optimal saving problem with a Shannon capacity constraint and non-quadratic utility. It derives an optimal ex post probability density for wealth in two leading cases (log and linear utility) and lays out a general approach for handling other cases numerically. It displays and discusses numerical solutions for other utility functions, and considers the feasibility of extending this paper's approaches to general non-LQ dynamic programming problems. The introduction of the paper discusses approaches that have been taken in the existing literature to applying Shannon capacity to economic modeling, making criticisms and suggesting promising directions for further progress.

I. INTRODUCTION

In a pair of previous papers that consider this topic(Sims, 2003, 1998)[1] I have argued for modeling the observed inertial reaction of economic agents to external information of all kinds as arising from an inability to attend to all the information available, and for treating that inability as arising from finite Shannon capacity. Shannon capacity is a measure of information flow rate that is inherently probabilistic. It uses the reduction in the **entropy** of a probability distribution as the measure of information flow. The entropy of a distribution is a global measure of the uncertainty implied by the distribution, relative to some base distribution. Because of this dependence on the base, the entropy of a distribution is not uniquely defined, but if we consider the joint distribution of two random vectors or variables, the expected reduction in entropy of one of the two achieved by observing

Date: March 14, 2006.

This research was supported in part by NSF grant SES-0350686. The paper was originally prepared for and presented at the Deutsche Bundesbank Spring Conference, May 2005.

[1] The earlier of these contains an appendix arguing that Shannon capacity makes sense as a model of inattention. The later one gives explicit solutions for some simple economic models with a linear-quadratic structure.

the other of the two, the **mutual information** implied by the joint distribution, is uniquely defined, independent of any base. This measure of mutual information can be derived from a few reasonable axioms, but it is pervasive less because of its axiomatic appeal than because has proved to be exactly the concept appropriate for studying information flows in physical communication channels. A Shannon "channel" is a set of possible inputs, a set of possible outputs, and a conditional distribution for outputs given inputs. From these elements, it is possible to calculate a tight upper bound for the mutual information between inputs and outputs, which is called the channel's **capacity**. It is the measure of information flow we use in characterizing modems or internet connections in bits per second or bytes per second. Shannon showed that no matter what we might wish to send through the channel, whether music, text, or spreadsheets, and no matter what the physical nature of the channel — wires, optical cables, radio transmission, or a messenger service — it is possible to send information through the channel at a rate arbitrarily close to capacity.

Economists, particularly macroeconomists, have recognized the need to account for the inertia in observed economic behavior and have modeled it with a variety of devices — menu costs, adjustment costs, information delay, implementation delay, etc. As my two earlier papers argued, these mechanisms can match the observed pattern — slow, smooth cross-variable responses, combined with less smooth idiosyncratic randomness — only by postulating elaborate inertial schemes that are both difficult to connect to observation or intuition and critically important in making model behavior realistic. One appeal of the rational inattention idea (that is, of modeling agents as finite-capacity channels) is that it can in principle explain the observed patterns of inertial and random behavior by a mechanism with many fewer free parameters. Another is that it fits well with intuition; most people every day encounter, or could very easily encounter, much more information that is in principle relevant to their economic behavior than they actually respond to. The notion that this is because their are limits to "attention", and that such limits might behave like finite Shannon capacity, is intuitively appealing.

II. Recent Developments in the literature

A number of recent papers in macroeconomics and finance have used information-theoretic ideas (Maćkowiak and Wiederholt, 2005; Luo, 2004; Mondria, 2005; Moscarini, 2004; Van Nieuwerburgh and Veldkamp, 2004a,b; Peng and Xiong, 2005; Peng, 2005). While these papers develop some valuable insights, it is worth noting that they have made assumptions, to allow tractable modeling, that are hard to defend and can lead to anomalous results. Some of these limitations are common to all or nearly all of the papers.

II.1. **Not allowing fully endogenous choice of the form of uncertainty.** It is central to the idea of modeling individuals as capacity-constrained that the nature, not just the quantity of their uncertainty about external signals (prices, income, wealth, asset yields, etc.) is subject to choice. The power of information theoretic ideas arises from the fact that the available joint stochastic processes for channel input and channel output are, to an arbitrarily good approximation, limited only by the capacity of the channel, not by its physical nature. In a model of an optimizing agent, the agents' objective function will therefore determine the stochastic process for the joint behavior of actions and external signals. The articles cited in the previous paragraph, with the partial exception of Luo's, postulate directly a simple parameteric form for this joint process, without deriving that form from the model's objective function.

To be more specific, the papers all assume Gaussian prior uncertainty about a state variable and Gaussian posterior (after information flow) uncertainty. Furthermore, in some cases they assume that the prior and posterior uncertainty is over a random vector and is i.i.d., either over the elements of the vector itself or over a set of factors that generate the distribution of the vector. (Luo and Mondria do consider endogenous choice of posterior covariance structure.) It is true that Gaussian posterior uncertainty can be shown to be optimal when the loss function is quadratic, but only Luo's paper considers cases of pure quadratic loss. Even if the loss function is quadratic, it is not generally optimal for a capacity-constrained agent to have i.i.d. posterior uncertainty across the same variables or factors that were a priori i.i.d. As we will see in some examples below, standard forms of utility functions in an economic model generate strongly non-Gaussian forms of optimal posterior uncertainty.

II.2. **Back-door information flows.** Several of the papers develop market equilibria, and to avoid complications assume that market prices are observed without error. But in these equilibria market prices are information-carrying random variables. Assuming they can be observed without error amounts to assuming unbounded information-processing capacity.[2] Counter-intuitive results can emerge when we assume perfect observation of prices combined with capacity-constrained observation of some other source of information.

II.3. **Distinguishing human information use, costly external information transmission and costly investigation.** The models in Sims (2003) and those presented

[2]An infinitely long sequence of digits can carry an infinite amount of information. Such a sequence, with a decimal point in front of it, is a real number. So if I can transmit an arbitrary real number without error in finite time, I have an infinite-capacity channel.

below in this paper are motivated by the idea that information that is freely available to an individual may not be used, because of the individual's limited information processing capacity. That capacity is unitary, allocatable to control many dimensions of uncertainty the individual faces. The "price" of this information is the shadow price of capacity in the individual's overall optimization problem.

Individuals or firms may also choose Shannon capacities of periodical subscriptions, telephone lines, internet connections, and other "wiring" that brings in information from the outside world. For a financial firm with a large staff constantly active in many markets, the wiring costs of information may indeed be more important, or at least comparable to, the costs of mapping information, once it is on the premises, into human action. However for most individuals, wiring costs are likely to be small relative to the costs associated with human information processing. In any case the costs of the two kind of information will generally be quite distinct, on a per-bit basis. One can't replace the human decision-making that links prices, incomes and wealth to real actions with a fiber-optic cable.

Both wiring and internal human information processing are reasonably measured in bits, with costs linear, or at least smooth, in bits. There is another kind of "information", however, whose cost is different, and probably usually not well measured in bits. In the stock market, an individual investor has a vast amount of information about individual stocks available at practically no or trivial cost, in newspapers and on the internet. It is likely that he does not use all this information, due to limited information processing capacity. But it is also possible to develop information through costly investigation — interviewing experts in a firm's technical area, conducting surveys of consumers to determine their reactions to the company's product, etc. The CEO of a drug company might contemplate approving a clinical trial to determine whether or not a new drug is an improvement on existing treatments, approving a focus group investigation of which of two packages is most preferred by consumers, or stepping outside to see if it's raining. Each of these three actions would (if the answers had 50-50 probability in advance) yield one bit of information. But it is no help to decision making to think of them as bits limited by a capacity constraint.

Several of the finance-oriented papers cited above consider at least some models in which uncertainty about an asset's yield is quantified as the standard deviation of its distribution, and information costs are quantified as bits, measured by reduction in the log of the standard deviation. But this is only appropriate if the information is thought of as freely available, with only wiring costs or human capacity costs preventing it from being known with certainty. In asset markets this is almost never the case. Sophisticated, continuously trading investors have uncertainty that is dominated by information that is not freely available, and less sophisticated investors, who do fail to use instantly all freely available information, do not have the option of reducing the log standard error of their uncertainty about yields to arbitrarily low levels according to a linear cost schedule.

While failure to make these distinctions does not necessarily make a model uninteresting, it can make a model's interpretation difficult. Especially in highly liquid financial markets, It is probably important to recognize that wiring capacity and human information-processing capacity have different costs. It is certainly important to to distinguish information about asset returns that is freely available but costly to act upon from information that can be obtained only through costly investigation.

III. Moving beyond the LQ Gaussian case

For some purposes linear-quadratic Gaussian (LQG) models may give reasonable approximations. Luo (2004) applies information-theoretic ideas to optimization problems with linearized first order conditions as are commonly used recently in macroeconomic modeling. The idea is that if uncertainty is fairly small, linear approximations to the model's FOC's may be quite accurate, so that the LQG framework remains an adequate approximation even with capacity constraints. While this is an idea worth pursuing, because it yields insights and is tractable, there is reason to worry about its range of applicability. If rational inattention is to explain much of observed inertia in behavior, people must be using a small part of their capacity to monitor economic variables. But in this case information-processing based uncertainty will be large, and this in itself will tend to undermine the accuracy of the local LQG approximation. Also, there are many interesting issues, like the interaction of finite capacity with the degree of risk aversion that is investigated below, that cannot be studied in an LQG framework.

In this section, therefore, we show that moving beyond the LQG framework is feasible. We consider several variations on a simple two-period saving problem. The problem is so simple that the information flows we will be looking at are unrealistically low. Nonetheless it is interesting to see that the model provides some insights into behavior, is computationally manageable, and suggests that a more interesting fully dynamic version might be feasible.

The problem is

$$\max_f \int_{0<c<w} \log\big(c\cdot(w-c)\big)\, f(c,w)\, dw\, dc \tag{1}$$

subject to

$$f(c,w) \geq 0 \tag{2}$$

$$\int_{0<c<w} f(w,c)\,dc = g(w) \tag{3}$$

$$\begin{aligned}\int_{0<c<w} \log\big(f(c,w)\big)\cdot f(c,w)\,dw\,dc \\ - \int_0^\infty \left(\log\left(\int_c^\infty f(c,w)\,dw\right)\cdot \int_c^\infty f(c,w)\,dw\right)\,dc \\ - \int_0^\infty \log\big(g(w)\big)\cdot g(w)\,dw \le \kappa\,. \end{aligned} \tag{4}$$

The expression (1) is a standard assertion that we are maximizing expected utility, where that is the sum of the expected utility of current consumption, $\log c$, and that of next period's consumption, $\log(w-c)$. (We could include a gross interest rate greater than one and a discount factor less than one without changing anything important.) What is unusual is that the "choice variable" with respect to which we maximize is not current consumption c, but the joint pdf of c with wealth w. The constraint (2) recognizes that $f = 0$ puts us at the boundary of feasible values for probability densities. The constraint (3) tells us that the marginal distribution of wealth is fixed, so all that is available for choice is $f(c,w)/g(w)$, the conditional pdf of c given w. The information constraint is (4). The last term in (4) is the entropy of the marginal distribution of w, the next to last term is the entropy of the marginal distribution of c, and their sum is what the entropy of c and w's joint distribution would be if they were independent. The first term is minus the entropy of the actual joint distribution determined by f. The the three terms together form the *mutual information* between c and w. This is also the expected reduction in the entropy of the w distribution from observing c, and also the expected reduction in the entropy of the c distribution from observing w.

The first order condition for the problem is

$$\begin{aligned}\log\big(c\cdot(w-c)\big) \\ = \lambda\left(1+\log f(c,w) - 1 - \log\left(\int_c^\infty f(c,w)\,dw\right)\right) + \mu(w) + \psi(c,w)\,. \end{aligned} \tag{5}$$

Here μ is the Lagrange multiplier on (3) and $\psi(c,w)$ is a stand-in for the fact that when $f = 0$, the FOC's do not have to hold. (Since at $f = 0$ we will have $\log f = -\infty$, no finite value of $\psi(c,w)$ makes the FOC hold when $\log(c\cdot(w-c))$ is finite, but the non-convexity of the constraint set means that solutions on the $f = 0$ boundary can nonetheless occur at such c, w values.) If we let $q(w \mid c)$ denote the conditional pdf of w given c, $\alpha = 1/\lambda$, and $\nu(w) = e^{-\alpha\mu(w)}$, this expression can, at points where $f > 0$, be rearranged as

$$q(w \mid c) = \nu(w)c^{\alpha}(w-c)^{\alpha}\,. \tag{6}$$

The function ν must, according to (6), make the integral of the right-hand side with respect to w one, regardless of the value of c. One ν that works (the only one?) is a ν proportional to $w^{-2\alpha-1}$. With this choice of ν, if we rewrite in terms of $v = w/c - 1$, the integral becomes

$$\int_c^{\infty} \nu_0 w^{-2\alpha-1} c^{\alpha}(w-c)^{\alpha}\,dw = \nu_0 \int_0^{\infty} (v+1)^{-2\alpha-1} c^{-2\alpha-1} c^{\alpha} v^{\alpha} c^{\alpha} c\,dv\,. \tag{7}$$

Since the terms in c cancel, the integral does not depend on c, and by choosing ν_0 properly we can make the integral one.

The form of the integrand in (7) is proportional to that of an $F(2\alpha+2, 2\alpha)$ density. Normalized to have constant spread, this does approach normality as $\alpha = 1/\lambda$ approaches infinity, i.e. as the shadow price of information in utility units approaches zero. The peak is at $(2 - 1/(\alpha-1))c$, the mean at $(2 + 1/(\alpha-1))c$, for $\alpha > 1$. The distribution is more tightly concentrated around $w = 2c$ the larger is $\alpha = 1/\lambda$. In other words, as the shadow price λ on the information constraint declines, we come closer and closer to the certainty solution.

Figure 1 shows a contour map of the pdf of $w \mid c$ for a case where $\lambda = .5$. Note that the conditional distribution of $w \mid c$ in this case is centered roughly at $2c$, as we expect. The solid line on the figure shows the value of c that would be chosen under certainty, $w/2$, and the dotted line shows $c = w$, which is the lower bound on w for a given c.

Figure 2 displays these conditional densities in a different way. Each line shows $q(w \mid c)$ for a different value of c between zero and .5. These are all of course truncated, scaled F densities. They have all been scaled to integrate to 1 over the $(0,1)$ interval, to make them comparable to numerically derived densities we will show below.

As is true for the LQ case, we find here that the form of the distribution for w conditional on available information at decision time is invariant to $g(w)$, the marginal pdf for w before information flow, so long as the density has full support. This is not to say that the conditional distribution itself is invariant to g. If g has high entropy, then with a given κ it will not be possible to reduce entropy much, λ will be large, and the α parameter that determines how close c is to $w/2$ will be small. But there is a single parameter, λ, that controls all the possible variation in the form of the distribution of $w \mid c$ when f has full support.

The first-order conditions need not hold at points where the joint pdf $f(c,w)$ is zero. Obviously if $g(w)$ is zero over some range, $f(c,w)$ must also be zero over the corresponding range of w values. Even where $g(w)$ has the whole non-negative

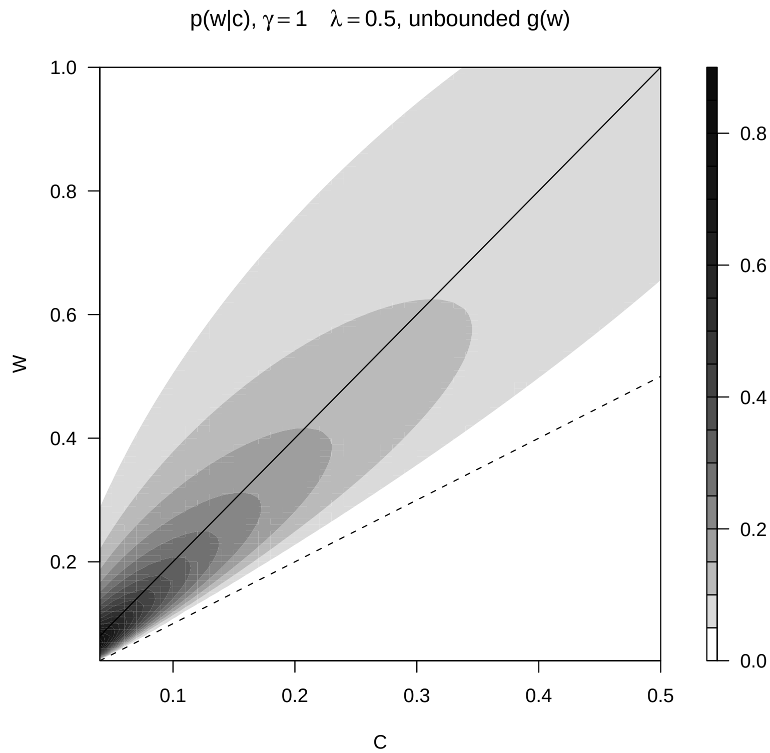

FIGURE 1.

real line as its support, it is possible for $p(c)$ to emerge as zero over some set of c values and indeed even possible for the marginal distribution of c to emerge as discrete. If the conditional distributions of c given w has, over any set of w's with non-zero probability, discrete weight on points that do not have discrete weight in the marginal distribution of c, this would imply an infinite information flow. But so long as any c values getting discrete weight in the conditional distributions also have discrete weight in the marginal, the information flow is finite. Of course at discrete values of c that have non-zero probability, the first-order conditions of the problem hold, so for these c's, as for any others with positive density value, we expect $q(w \mid c)$ to take on the form we derived above, over its support. It is possible, though, for $q(w \mid c)$ to have less then full support. This obviously allows deviation from the F distribution we derived above, but of a particular form, since it must have the shape of an F density over its support.

While the distribution of $w \mid c$ is easy to characterize here, it is not easy (for me, anyway) to characterize the distribution of $c \mid w$, even in the case where the

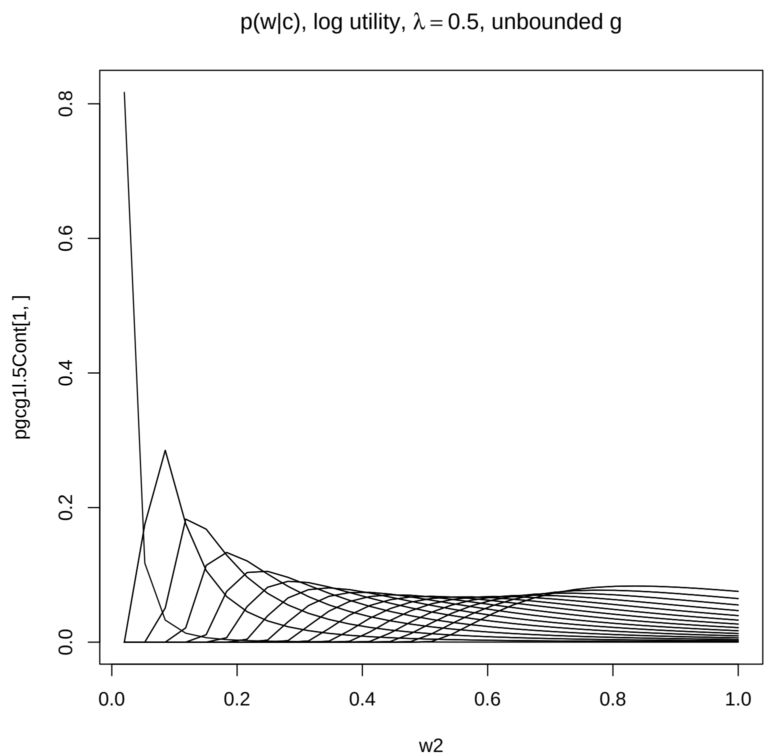

FIGURE 2.

marginal on w is assumed to have the same scaled-F form as the post-observation distribution for savings, or to find a form for the marginal of c that, together with the known form for $w \mid c$, implies that the marginal for w is a scaled F. However this is just a very simple example. Fully dynamic models are likely to generate distributions complicated enough to require numerical methods for solution in any case.

We can see an analytic solution for one other simple special case: where the utility function is linear and the $c < w$ constraint is maintained.[3] If the utility function is undiscounted, so $U(c, w) = c + w - c = w$, the problem has the trivial solution $c = 0$, with no information at all used. The problem is a little more interesting if utility is discounted, so $U = c + \beta(w - c)$ with $0 < \beta < 1$. For any U, the FOC's take the same form as (5), but with $U(c, w)$ replacing the log function on the left of

[3]Without the $c < w$ constraint, and with discounting, agents who can borrow at zero interest will obviously push c to infinity.

the equality. For this linear case, the analog of (6) is

$$q(w \mid c) = v(w)e^{\alpha(\beta w + (1-\beta)c)} \tag{8}$$

By choosing $v(w)$ proportional to $e^{-\alpha w}$, we get this in a form that, when integrated from c to ∞ w.r.t. w, gives a constant value. The implied form for the conditional pdf of $u = w - c$ given c is $\alpha(1-\beta)e^{-\alpha(1-\beta)u}$. Here we can note that as α increases (so information is flowing more freely) the solution converges toward $c = w$, which is the optimum without uncertainty. It is also interesting that c and w are implied to be independent conditional on any rectangle inside the $w > c$ region. These risk-neutral agents waste no capacity on matching c to w itself, except as it contributes toward knowing where the $w = c$ boundary is.

In the model with log utility, capacity-constrained agents have expected wealth, given their consumption, that exceeds the level corresponding to the deterministic solution $w = 2c$. The higher are information costs (the lower is capacity), the longer is the tail on the $w \mid c$ distribution and the larger the excess $E[w \mid c] - 2c$. There is an effect that seems to go in the opposite direction, of course: the mode of the $w \mid c$ distribution falls further *below* the deterministic value as information costs increase. However when data are aggregated across many individuals in different circumstances, we would expect the expectation result to dominate. We thus see a "precautionary savings due to information costs" effect.

But notice that the model with linear utility produces the opposite result. These risk-neutral agents who discount the future, while facing a gross rate of return of 1, are constrained from consuming all their wealth in the first period only by their uncertainty about what what that total wealth is. Relaxing their capacity constraint produces less saving.

With quadratic utility, the left-hand side of (5) is quadratic. Normalizing the utility function to $U(c,w) = c - \frac{1}{2}c^2 + (w-c) - \frac{1}{2}(w-c)^2$ leads to the analog of (6) as

$$q(w \mid c) = v(w)e^{\alpha(w-(c^2+(w-c)^2)/2)}\,. \tag{9}$$

If we drop the $c < w$ constraint and also the $c > 0$ constraint, the right-hand-side of (9) as a function of w is proportional to a Gaussian pdf with variance $1/\alpha$, with only the mean of the distribution dependent on c. Hence we can make the right-hand-side's integral one by choosing $v(w)$ to be constant. It is then easy to verify that if the exogenously specified marginal pdf for w, $g(w)$, is Gaussian, the joint pdf for c and w is Gaussian. Observe that whatever $g(w)$ we start with, so long as it allows a solution with $f > 0$ everywhere, the conditional distribution of w is Gaussian. Thus if this problem were part of a recursive scheme, all the joint distributions of successive c's and w's after the first period would be Gaussian.

However, this result depends crucially on there being no $c < w$ or $c > 0$ restriction. With these restrictions, despite the form of (9), the dependence of limits of integration on c will require at least a non-constant $v(w)$, and possibly some regions of $f = 0$.

So we can conclude from these examples:

- A hard budget constraint is not incompatible with a finite rate of information flow. The conditional distribution of $c \mid w$ can be confined to the $(0, w)$ interval, even though observation of neither c nor w ever gives perfect information about the other variable. An agent behaving this way would be making decisions that only imperfectly determine c, based on his imperfect knowledge of w. For example, writing checks or using credit cards and occasionally finding that the account is overdrawn, or getting to the checkout counter of the grocery store and realizing he will have to put a few things back, or buying $10 worth of gasoline without figuring out in advance how many gallons that will be.
- Uncertainty arising from information processing can easily be quite non-normal, even when exogenous shocks are small. In this example, there are no exogenous shocks. Normality is a good approximation only when the information constraint is not having a strong effect.
- A capacity constraint can have powerful implications for savings behavior. This accords with the facts that most people only vaguely aware of their net worth, are little-influenced in their current behavior (at least if under 50) by the status of their retirement account, and can be induced to make large changes in savings behavior by minor "informational" changes, like changes in default options on retirement plans.

IV. Some models that require a computational approach

I have no recipe for exhaustively identifying cases like log utility, linear utility, and quadratic utility without borrowing constraints, in which an analytic solution for $q(w \mid c)$ is obtainable. Indeed I have the impression that such cases are very rare. So it is worthwhile to look at some examples of commonly used $U(x, y)$ functions and see how hard it is to compute solutions.

For an $f > 0$ solution, the first-order conditions and the constraint that the marginal pdf for w be the given $g(w)$ lead to the pair of equations

$$\int e^{\alpha U(c,w)} v(w) dw = 1\,, \quad \text{all } c, \tag{10}$$

$$\int h(c) e^{\alpha U(c,w)} v(w) dc = g(w)\,, \quad \text{all } w\,, \tag{11}$$

which have to be solved for v and h, where h is the marginal pdf of c. This is a recursive linear system. It looks like we could discretize it, solve the resulting simple linear system from (10) for v, then use those results in (11) to create another simple linear equation system to solve for h. This approach does not work.

Each of these equations is what is known as a Fredholm integral equation of type 1, which are notoriously ill-conditioned except in special cases. In other words, a v that makes the norm of the vector of discrepancies between right and left hand sides of (10) nearly zero, can differ from the true solution in v-space by a large amount. Furthemore, it is not going to be uncommon for there to be regions of c, w space with $f(c,w) = 0$ in the solution. If one knew where these were, (10-11) could be used on the remaining c, w values. But in general we will not know where they are, and searching over all the possible combinations of such regions is prohibitively complicated.

An approach that I have found to work is simply to discretize f itself and maximize the Lagrangian

$$\int U(c,w) f(c,w)\, dc\, dw - \lambda H(W,C) - \mu(w) \left(\int f(c,w)\, dc - g(w) \right) . \qquad (12)$$

with λ fixed at some positive number. Here $H(W,C)$ is the mutual information between w and c in their joint distribution, the same object that appears on the left-hand side of the information constraint (4). This can work because points at which $f = 0$ simply drop out of both the information constraint and the expected utility. I have imposed $f > 0$ by maximizing over $\log f$ as the parameter vector. This means of course that the parameters corresponding to $f(c,w) = 0$ values are ill-determined, but gradient-based search methods (at least my own, `csminwel.R`, which is what I used) still perform well, converging nicely for the $f(c,w) > 0$ values and leaving $\log f$ extremely negative at points where clearly $f(c,w) = 0$.

The discretized solutions below all are based on using an equi-spaced grid with 16 c values ranging from .01 to .5 and 31 w values ranging from .02 to 1. The marginal pdf g for w is given in each case as $g(w) = 2w$. There are then 31 adding-up constraints, and $8 \times 15 = 120$ zero constraints on points where $c > w$, leaving $31 \times 16 - 120 - 31 = 345$ free parameters. The specific normalization I used was an unconstrained 15×31 matrix θ of parameters, with the entry f_{ij} of the discretized f determined as $g_j \exp\big(\theta_{i-1,j}\big)/(1+\bar{\theta})\big)$ for $i > 1$ and as $g_j/(1+\bar{\theta})$ for $i = 1$, and with $\bar{\theta} = \sum_i \exp(\theta_{ij})$.

Though for some problems 345 parameters would be so many as to raise computational difficulties, here there seems to be no trouble with them. Using interpreted (in the R language[4]) code and numerical derivatives, convergence is achieved in

[4]Available at `http://lib.stat.cmu.edu/R/CRAN/`

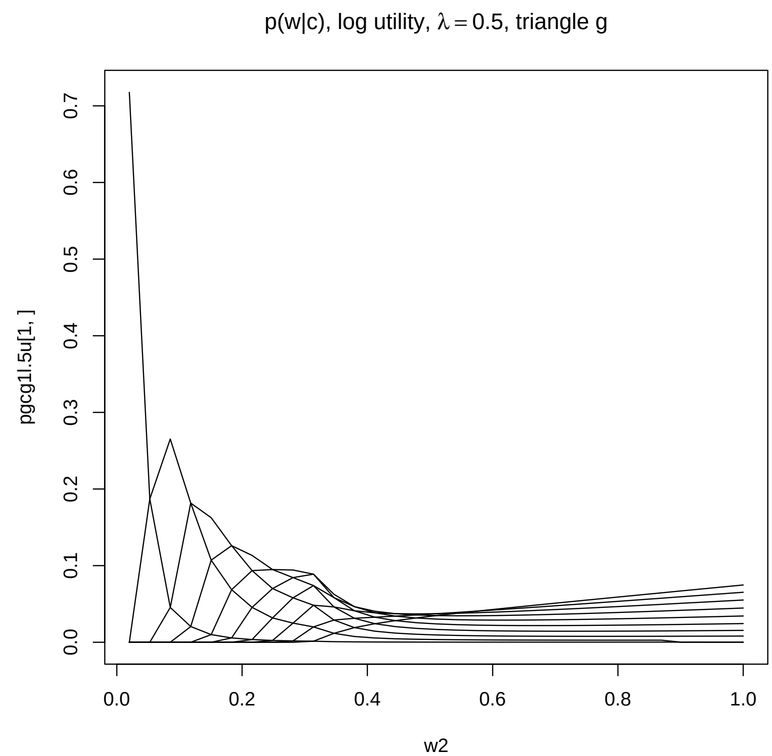

FIGURE 3.

about 450 iterations, with each iteration taking about 1.5 seconds (making the whole computation take about 11 minutes) on a 3GHz Pentium 4 running Linux. They could be made much faster using analytic derivatives, which would not be hard to program, and probably also by using compiled code. A more sophisticated algorithm could also help. We know the solution should be smooth on its support and should have a density with respect to some product measure on c, w space. Parameterizing the solution to reflect this knowledge should enable accurate approximations with many fewer free parameters.

Figure 3 shows the same kind of plot as Figure 2 — $q(w \mid c)$ densities with various values of c, log utility, and $\lambda = .5$. The difference is that here, instead of showing the theoretical densities for the case where the $q's$ (and hence necessarily also g) have unbounded support, we show the numerically computed densities for our case of a linearly increasing density with support $[0, 1]$. The lines on the two plots should not match exactly, because one shows a discretized approximation to the solution for continuously distributed c and w, while the other shows an exact solution for discretely distributed c and w. Nonetheless these conditional pdf's do

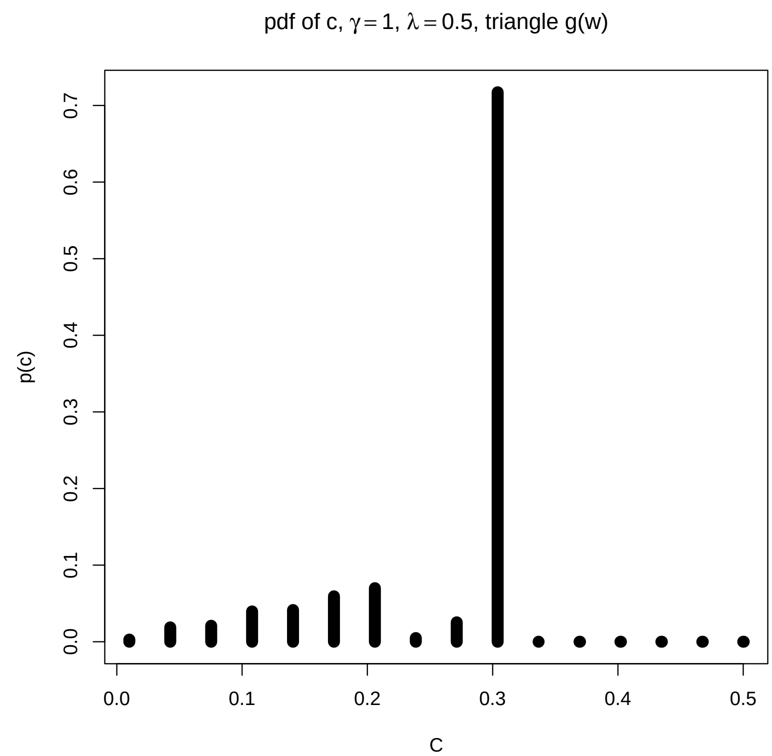

FIGURE 4.

match up, roughly. Note that in Figure 3 there are fewer lines shown, because the discreteness and the bounded support for g mean that c does not have full support even in $[0, .5]$, as can be seen from Figure 4. In fact, with this degree of information constraint, the solution sets c to .4 with probability .72.

To see the effects of varying degrees of risk aversion, consider calculations for the same triangular $g(w)$, but with utility given by the CRRA form

$$U(c, w) = \frac{c^{1-\gamma} + (w - c)^{1-\gamma}}{1 - \gamma}\,. \tag{13}$$

The values of λ in the two examples we consider have been adjusted so that the information flow (.88 bits and .85 bits), is about the same for the two cases ($\lambda = .5$ and $\lambda = 2$, respectively) considered, so the differences in results are attributable to the differences in the risk aversion parameters. Of course if the model were formulated with an actual opportunity cost of information, in consumption goods units, the solution with lower risk aversion would most likely imply a choice of lower information flow.

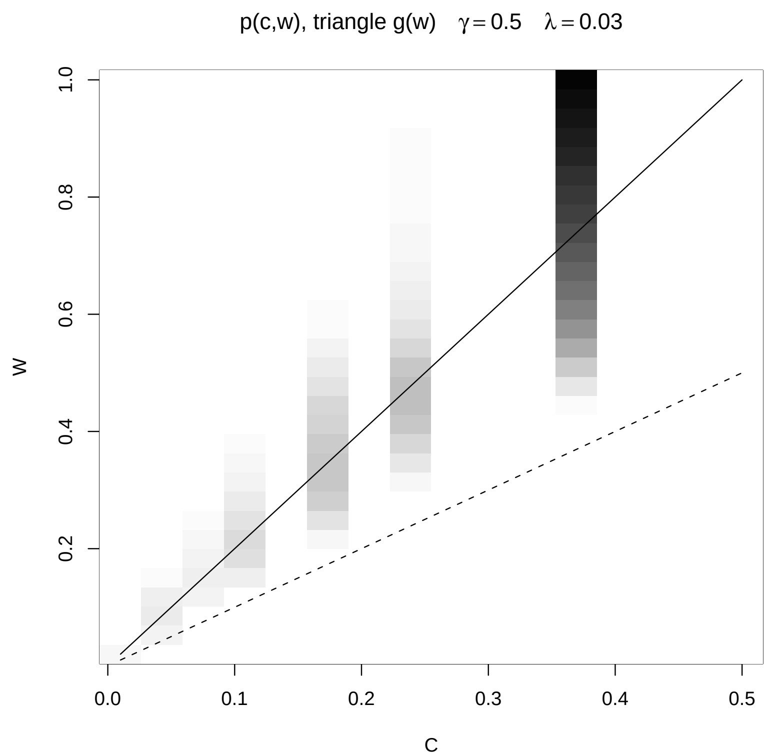

FIGURE 5.

Shaded plots of the joint densities of c and w are shown in Figures 5 and 6. Both solutions show discretization, with several values of c receiving zero probability. The $\gamma = .5$ solution puts probability .70 on the highest value of c, which is .37. The $\gamma = 2$ solution also puts probability .70 on its highest c value, but the value is slightly lower — .34. The low risk-aversion solution shows a more discretized distribution of c at lower values of w, and puts substantially more probability on c values close to the $c = .5w$ line. This is as would be expected. The $\gamma = 2$ utility function goes to minus infinity as $w - c \to 0$ or $c \to 0$. It therefore makes it worthwhile to be well informed about low wealth values, so that c is not forced too close to zero, and also to keep the probability of choosing c close to w (and hence $w - c$ small) low.

With these non-log utility functions, we expect to see deviations from the F form of $q(w \mid c)$. Plots analogous to Figures 2 and 3 for these two non-log utility examples are shown in Figures 5 and 8. The high-risk-aversion plot shows narrow finite support for $q(w \mid c)$ at low values of c, and then very flat tails on the pdf's as we move to higher c values. The lower-risk-aversion plot shows more of a right tail

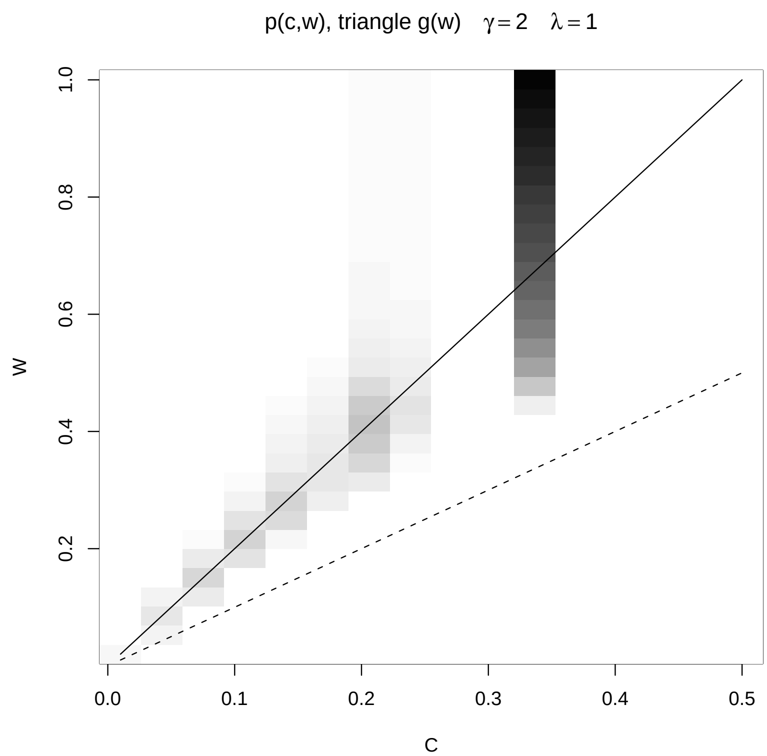

FIGURE 6.

and less narrow support at the low c values, and thin tails at higher c values except for the highest, which in both plots is the one with the most weight. These results make sense, with the higher risk aversion solution trading higher probability of choosing a modest value of c when wealth is in fact high for a lower probability of a mismatch between c and w when wealth is low.

The fact that discreteness in the distribution of c and finite support, varying with c, in $q(w \mid c)$ show up in both these standard cases, despite a continuous marginal distribution for w, implies that any numerical approach to solution of rational inattention models must allow for these possibilities. They are not an artifact of the triangular $g(w)$. That $g(w)$ was chosen to bring out the differences with degree of risk aversion, but the discreteness of the c distribution showed up with every g I considered, including standard Beta, F, and Gamma pdf's that peaked below .3 and were small or zero at $w = 1$.

While the discreteness will be a challenge for programmers, it may help the theory rationalize observed patterns of behavior. Actual choices by individuals in response to external information often do seem to have a discrete character.

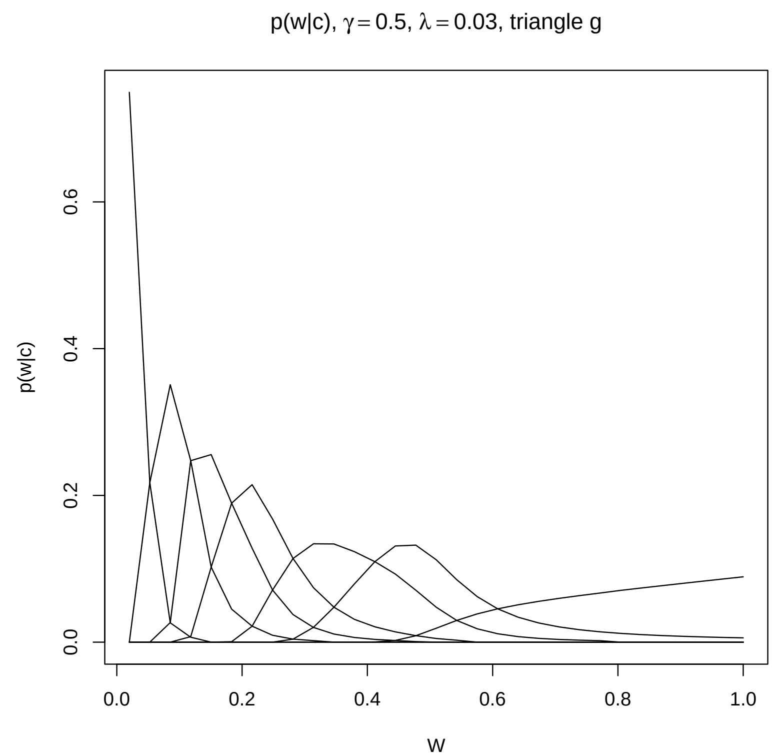

FIGURE 7.

V. OTHER POINTS ON THE RESEARCH FRONTIER

V.1. **On to a fully dynamic non-LQG model.** Here is the Bellman equation for a dynamic programming problem with Shannon capacity as a constraint, the current pdf g of w as the state variable, and $f(c,w)$ as the control:

$$V(g) = \max_{f(\cdot,\cdot)} \int U(c) f(c,w)\, dc\, dw + \beta \int V\left(\int h(\cdot; c, w) f(w \mid c)\, dw \right) f(c,w)\, dc\, dw \tag{14}$$

subject to

$$\int f(c,w)\, dc = g(w), \text{ all } w \tag{15}$$

$$f(c,w) \geq 0, \text{ all } c, w \tag{16}$$

$$H(C,W) \leq \kappa\,. \tag{17}$$

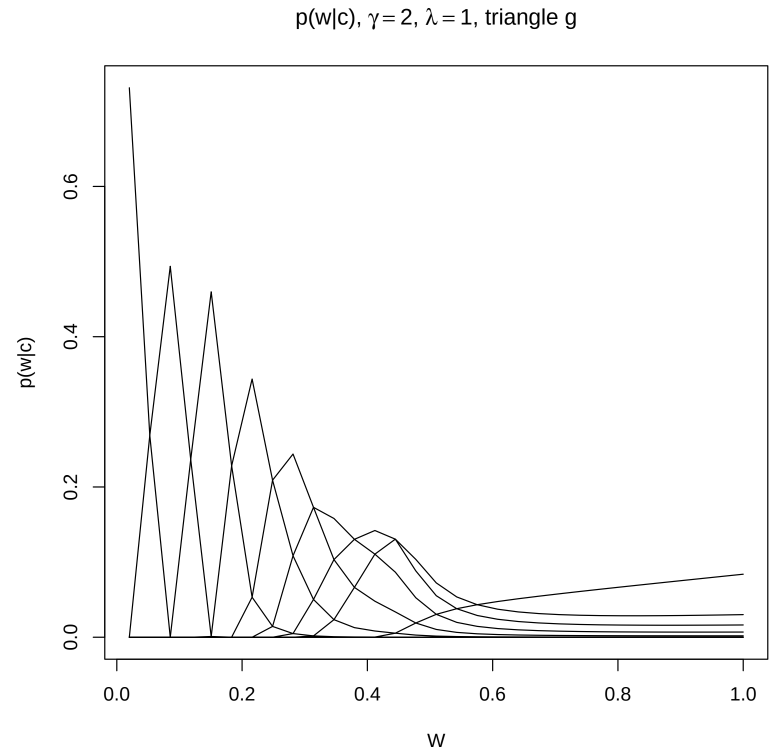

FIGURE 8.

The function h maps the current c, w pair into a conditional density for next period's w. In usual models, it is specified indirectly in the form of an equation like $w_t = \phi(w_{t-1}, c_{t-1}, \varepsilon_t)$ together with a specification that ε_t has a certain pdf and is independent of w_{t-1} and c_{t-1}. The constraint connecting $f(c, w)$ to $f(c \mid w)$ has been left implicit. This problem is in the form of a standard dynamic programming problem, except that the state and control variables are both in principle infinite-dimensional. But, extending the approach taken above to a two-period problem, I believe computing solutions to such problems should be feasible. Economists are already succeeding in calculating solutions to equilibrium models with infinite-dimensional state spaces.

Note the occurrence of $f(w \mid c)$ in the argument of the value function on the right-hand side in (14). This reflects the fact that the agent must allow some "noise" to affect the choice of c in the current period, but can use the noisy observation that entered determination of c to update beliefs about next period's w.

V.2. **Rational inattention models of general equilibrium.** The work cited earlier by van Niewerburgh and Veldkamp and by Mondria includes calculation of market equilibrium with capacity constraints. As already noted in section II, though, these models assume costless and perfect observation of market prices, which is both contrary to the notion of finite-capacity agents and a source of anomalous results. This modeling choice by these authors is not an easily corrected oversight. Modeling a market equilibrium with agents who do not know exactly what prices are requires being explicit about aspects of market microstructure that are not standard parts of the economic theory toolbox and about which we have few stylized facts or modeling conventions to guide us.

A model populated by capacity-constrained agents will not simply balance supply and demand via a price mechanism. Agents will not have perfect knowledge of prices, indeed may have only a very rough idea of what they are, as they take decisions that affect economic exchange and production. Inventories, retailers, wholesalers, demand deposits, cash, and credit cards, are all devices that allow agents to make transactions in which quantities and price are known and chosen only approximately. Few of our models have explicit roles for retailers and wholesalers, our models of inventory behavior are only modestly successful, and microfounded models of money that connect to data are non-existent. The idea of Shannon capacity may be of some help in modeling these phenomena, but they are inherently difficult, long-standing problems, so realistic general equilibrium models with capacity-constrained agents may not emerge for some time.

V.3. **Macro modeling.** Since a standard approach to general equilibrium modeling with rational inattention will not emerge soon, applying rational attention to the representative agent equilibrium models that now constitute mainstream macro will also take some time. In the meantime, though, we can see even from simple linear-quadratic examples that there are implications for current modeling practice. In the linear-quadratic framework, rational inattention behavior is a constrained special case of the behavior of an agent who observes state variables with error. While there are some examples of such models in the macroeconomic literature (Lucas, 1973; Woodford, 2001), the rational inattention idea should encourage us to pay more attention to such models. The objection that there is no physical interpretation for the observation error such models postulate is answered by the rational inattention framework, and the RI framework gives us some guidance as to reasonable properties for the observation error, even when we cannot derive it analytically.

V.4. **Public and private information models.** Recently, following the paper by Morris and Shin (2002), there have been a number of papers considering models with public and private information (Hellwig, 2004; Angeletos and Pavan, 2004). This work raises interesting questions and arrives at conflicting conclusions about the value of "transparency", depending on assumptions about externalities that are difficult to calibrate against an actual economy. But for all the diversity in the conclusions from this literature, it is all formulated on the assumption that there are private information sources with exogenously given attributes and a public information source whose stochastic character we can imagine controlling. From a rational inattention perspective, this is a strange setup. Capacity-limited agents will act as if observing the state of the economy with error even if some public authority announces it exactly. The amount of the error will depend both on the stochastic properties of the state itself and of any noise in public signals about it. If private signals carry a lot of information about privately important variables, information they contain about an aggregate state may be ignored or reacted to very slowly and erraticly[5]. Before these abstract models are applied to policy debates about transparency in monetary policy, they need to be tied closely enough to real economies that we can judge which of their conflicting conclusions might be correct, and this should include consideration of how their conclusions are affected by rational inattention.

VI. Implications for monetary policy

Rational inattention may have far-reaching implications for macroeconomics and monetary policy generally, once its implications are fully worked out. In the meantime, though, it may shed some light on transparency in monetary policy. For a capacity-limited agent, it is necessary to take actions that respond to the true state of the economy at a low information transmission rate. This means that reactions to the state are either delayed and smoothed, with added idiosyncratic error, or they are discretized and randomly timed. A central bank may provide the public with a heavily filtered view of its actions or judgments, on the assumption that the public cannot take in the full detail and complexity of its actions and the thinking behind them. In the US, this takes the form of discretized, somewhat randomly timed policy actions (changes in the Federal Funds rate) together with a brief paragraph rationalizing the action. The paragraph occasionally changes the wording of a phrase or two, and the market often reacts strongly to these changes. If enough market participants take this simple, discrete sequence of changes in wording as a free, low-bit-rate summary of an important state variable, it is unsurprising that markets respond discretely to these changes in wording. Even market specialists,

[5]See Maćkowiak and Wiederholt (2005) for a model with local and aggregate signals related to pricing decisions and responded to with capacity constraints.

who know the true state with high precision, will pay attention to the noise in the Fed signal because of its effects on other agents.

If this is what is going on currently, what would be the effect of the Fed's issuing a much more detailed inflation report, like those issued periodically by inflation-targeting banks? A naive extrapolation might suggest that since the Fed now provides very little information, which generates overreaction, overreaction would be much worse if the Fed proved much more information. But this is unlikely. Market participants who need a low-bit-rate summary of the state (of the economy, or of the Fed's policy stance) would still look for it, but there would not be a unique low-bit-rate summary in the Fed's own statements. Market participants who need to devote most of their attention elsewhere will still respond noisily and with delay to Fed statements, but the noise will come from other sources, probably many other sources, instead of mainly from the Fed's own efforts to provide a filtered signal. If there are enough other semi-public filterers of monetary news (TV, newspapers, investment clubs, lunchtable conversation), the signal processing noise in them may partially cancel out at the aggregate level. But even if not, the Fed would no longer be itself responsible for generating unnecessary market fluctuations.

There are other, in my view even stronger, arguments for transparency in monetary policy. But a rational inattention perspective does help us understand why it can be that economies where central banks publish detailed inflation reports do not seem to have as much of a problem with overreaction to those reports as the Fed does with overreaction to its single paragraphs.

VII. Conclusion

Abstract and single-agent models incorporating rational inattention are already providing us with some useful insights. There is still a long and interesting road ahead, though, before we can build models incorporating these insights that can be matched to observed data, either at the macro or micro level.

References

ANGELETOS, G.-M., AND A. PAVAN (2004): "Transparency of Information and Coordination in Economies with Investment Complementarities," *American Economic Review*, 94.

HELLWIG, C. (2004): "Heterogeneous Information and the Benefits of Transparency," Discussion paper, UCLA.

LUCAS, ROBERT E., J. (1973): "Some International Evidence on Output-Inflation Tradeoffs," *American Economic Review*, 63(3), 326–334.

LUO, Y. (2004): "Consumption Dynamics, Asset Pricing, and Welfare Effects under Information Processing Constraints," Discussion paper, Princeton University.

MAĆKOWIAK, B., AND M. WIEDERHOLT (2005): "Optimal Sticky Prices under Rational Inattention," Discussion paper, Humboldt University Berlin.

MONDRIA, J. (2005): "Financial Contagion and Attention Allocation," Discussion paper, Princeton University.

MORRIS, S., AND H. S. SHIN (2002): "The Social Value of Public Information," *American Economic Review*, 92, 1521–1534.

MOSCARINI, G. (2004): "Limited Information Capacity as a Source of Inertia," *Journal of Economic Dynamics and Control*, pp. 2003–2035.

PENG, L. (2005): "Learning with Information Capacity Constraints," *Journal of Financial and Quantitative Analysis*, 40(2), 307–329.

PENG, L., AND W. XIONG (2005): "Investor Attention, Overconfidence and Category Learning," Discussion paper, Princeton University.

SIMS, C. A. (1998): "Stickiness," *Carnegie-rochester Conference Series On Public Policy*, 49(1), 317–356.

——— (2003): "Implications of Rational Inattention," *Journal of Monetary Economics*, 50(3), 665–690.

VAN NIEUWERBURGH, S., AND L. VELDKAMP (2004a): "Information Acquisition and Portfolio Under-Diversication," Discussion paper, Stern School of Business, NYU.

——— (2004b): "Information Immobility and the Home Bias Puzzle," Discussion paper, Stern School of Business, NYU.

WOODFORD, M. (2001): "Imperfect Common Knowledge and the Effects of Monetary Policy," Discussion paper, Princeton University, `http://www.princeton.edu/~woodford/`.

DEPARTMENT OF ECONOMICS, PRINCETON UNIVERSITY
E-mail address: sims@princeton.edu